DOCUMENT RESUME

ED 110 513

TM 004 786

AUTHOR TITLE Niedermeyer, Fred C.; Sullivan, Howard J.

Development of Teacher-Administered Tests for the

SWRL Reading Programs.

INSTITUTION

Southwest Regional Laboratory for Educational Research and Development, Los Alamitos, Calif.

REPORT NO SWRL-TM-3-71-05

PUB DATE

16 Aug 71

25p.

EDRS PRICE DESCRIPTORS

MF-\$0.76 HC-\$1.58 PLUS POSTAGE

Criterion Referenced Tests; Elementary Education; Grade 1; *Individual Tests; Kindergarten; Multiple Choice Tests; *Predictive Ability (Testing); Reading

Diagnosis; *Reading Programs; *Reading Tests;

Response Style (Tests); Statistical Analysis; Teacher

Role; Test Construction; *Testing

ABSTRACT

To investigate the type of classroom testing format most appropriate for the Southwest Regional Laboratory (SWRL) Mod 2 Reading Program, three types of teacher-administered tests for the SWRL Second-Year Communication Skills Program were developed and tried out during the 1970-71 school year. The tests were administered by the classroom teacher as Criterion Exercises following each unit of instruction. Two of the tests were group-administered and had a multiple-choice format, with one type consisting of three-choice items and the other four-choice items. The third type of test was individually-administered, constructed-response format. Loth the individually-administered, constructed-response tests and the four-choice group-administered tests predicted well to an end-of-year criterion test, whereas the three-choice tests did not. Based on pupil-performance data and teacher reactions, the individually administered tests appear to be most appropriate for use as Criterion Exercises in the Hod 2 program. (Author/RC)



SOUTHWEST REGIONAL LABORATORY TECHNICAL MEMORANDUM

DATE:

August 16, 1971

NO:

TM-3-71-05

SCOPE OF INTEREST NOTICE

The ERIC Facility has assigned this document for processing

TW In our judgement, this document

in our judgement, in the clearing-houses noted to the right. Index-ing should reflect their special points of view.

PERMISSION TO REPRODUCE THIS COPY. RIGHTEO MATERIAL HAS BEEN GRANTED BY

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL IN-STITUTE OF EDUCATION FURTHER REPRO-DUCTION OUTSIDE THE ERIC SYSTEM RE-QUIRES PERMISSION OF THE COPYRIGHT OWNER

US DEPARTMENT OF HEALTH. EDUCATION & WELFARE

EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
HIS DOCUMENT HAS BEEN REPRO
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN
ATING IT POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRE
SENT OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY

DEVELOPMENT OF TEACHER-ADMINISTERED TESTS FOR THE SWRL

READING PROGRAMS

AUTHORS: Fred C. Niedermeyer and Howard J. Sullivan

ABSTRACT

To investigate the type of classroom testing format most appropriate for the SWRL Mod 2 Reading Program, three types of teacher-administered tests for the SWRL Second-Year Communication Skills Program were developed and tried out during the 1970-71 school year. The tests were administered by the classroom teacher as Criterion Exercises following each unit of instruction. Two of the tests were group-administered and had a multiplechoice format, with one type consisting of three-choice items and the other four-choice items. The third type of test was individuallyadministered, constructed-response format. Both the individuallyadministered, constructed-response tests and the four-choice groupadministered tests predicted well to an end-of-year criterion test, whereas the three-choice tests did not. Based on pupil-performance data and teacher reactions, the individually administered tests appear to be most appropriate for use as Criterion Exercises in the Mod 2 program.





DEVELOPMENT OF TEACHER-ADMINISTERED TESTS FOR THE SWRL READING PROGRAMS

This report describes the procedures used to develop a classroom testing format most appropriate for the SWRL Reading Programs. Some of the procedures utilized and knowledge gained may be generalizable to other objectives-based, teacher-administered instructional programs-particularly those concerned with the teaching of reading.

PROBLEM

In the SWRL Reading Programs, short criterion tests (called Criterion Exercises) are administered at two- to three-week intervals throughout the year to tell the teacher which children are achieving the program outcomes and which children need additional practice and remediation. Up through the 1969-70 school year, these tests had always been of a group-administered, selected-response format, with children marking their answers after choosing one of three choices for each item.

For some time, however, it had been suspected that these three-choice, selected-response Criterion Exercises (CEs) did not adequately identify those pupils needing remediation, i.e., the tests were too easy. Teachers contended that some children would score quite high on the CEs but still could not "read." To determine the extent to which this was true, a constructed-response test (where the child actually reads rather than selects the word or letter-sound) was individually administered to a random sample of 159 children at the end of the 1968-69 school year in 20 classes using SWRL's Kindergarten

Reading Program. Each child's score on this end-of-year mastery test was then paired with his average score over all of the CEs taken during the year. These pairs were plotted on a scattergram and correlation and regression analyses performed. This scattergram appears as Figure 1.

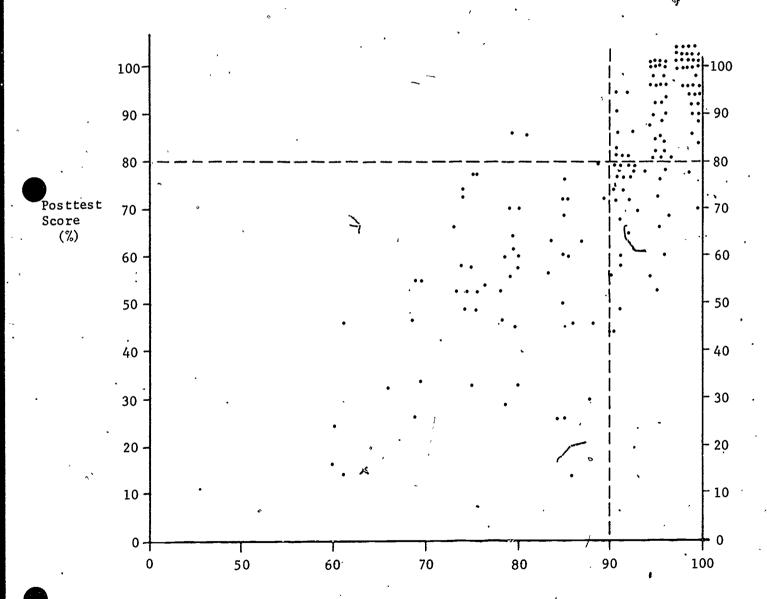
From Figure 1 two things may be noted. First, the relationship between the average CE scores and the scores on the end-of-year mastery test was, as might be expected, quite high (r=.80). More important, however, was the fact that a child had to obtain a very high average on the CEs (almost 95%) before one could predict that he would score at least 80% (a criterion of mastery established by SWRL) on the end-of-year test. In fact, 32 of the 159 children (20%) averaged 90% or better on the CEs but still scored below the 80% mastery criterion on the end-of-year posttest. Thus, not only were the CEs giving the teachers a false indication of achievement for these children during the year, but the high scores on the CEs prevented these children from being assigned additional instruction when, in fact, remediation was required.

The purpose of the present study was to develop and tryout other types of CEs--tests that would hopefully be more accurate indicators of how well a child would do at the end of the program. To this end, two types of CEs were developed in addition to the already existing three-choice, selected-response tests. For one type, a fourth distractor was added to each item of the existing three-choice CEs, thus creating a four-choice, selected-response CE. This fourth distractor was selected in such a way to make the item more difficult.



Figure 1. Scattergram showing relationship between the average CE score and the posttest score for 159 pupils in the SWRL Kindergarten Reading Program.

r=.80 p<.01



Average CE Score (%)

Finally, a different type of CE alrogether was developed. This was an individually-administered, constructed-response CE, which was similar in format to the end-of-year posttest. This type of CE had not been developed previously because it was felt that the group-administered, selected-response exercises were faster and easier to administer.

The procedures used and results obtained when developing and trying out these various types of CEs with the SWRL First-Grade Reading Program are described in the remainder of this report.

DESCRIPTION OF TESTS

This section describés each of the three types of tests developed and tried out during this study--the regular three-choice, selected-response Criterion Exercises (3-choice CEs); the newly-developed, four-choice, selected-response Criterion Exercises (4-choice CEs) and the newly-developed, individually-administered, constructed-response Criterion Exercises (constructed-response CEs).

3-choice CEs. The 3-choice CEs were already being utilized in the SWRL reading programs prior to this study. In the SWRL First-Grade Reading Program, there were 14 of these tests, one for each of the 14 two-week instructional units comprising the program. Each CE consisted of 24 items, with eight items for each of the three program outcomes-words, word elements and word-attack. The child's response booklet was four pages long with six items on a page. (See Figure 2 for a typical page from one of these CEs.) The first two rows of each page tested Outcome 1 (Words), the middle two rows always tested Outcome 2 (Word-Attack).



Figure 2. Sample page from one of the 3-choice CEs

1	l've		fly		we're	
2	very		blue		by.	•
3	У		x		v	. :
4	id		ix	¢	ick	v
5	· six	•	fix .		mix	
6	fix๋		dig		fig	

The teacher's script for the page shown in Figure 2 would read as follows:

Row 1: Mark the word I've.
Row 2: Mark the word blue.
Row 3: Mark vvv.
Row 4: Mark ick.
Row 5: Mark six.
Row 6: Mark fig.

4-choice CEs. The 4-choice CEs were developed directly from the existing 3-choice CEs by simply adding a fourth distractor to each item. The fourth distractor was generated in such a way as to make the item more difficult. For word recognition (Outcome 1, Rows 1 and 2), this usually meant adding another available word that possessed some of the same letters as the test word. For word elements of letter-sound correspondences (Outcome 2, Rows 3 and 4), a letter or letters were selected that were similar in construction to the tested element. word-attack (Outcome 3, Rows 5 and 6), the fourth distractor was generated wherever possible by finding a word with the same initial and ending letters as the test word, but with a different medial vowel. The previous 3-choice CEs had been constructed such that, if possible, one of the two distractors had the same beginning sound as the test word and the other the same ending sound. (See Item 6 in Figure 2.) For example, when testing the word_fig, the 3-choice exercises already had the distractors fix and dig. By adding fog as a fourth distractor, it seemed likely that this would make it harder for the child to figure out the correct answer by eliminating distractors, and would force him to simply read the four words to find the test word. It had been felt that many children were able to "break the code" on the 3-choice CEs by

eliminating distractors with different beginning or ending sounds than the word given orally by the teacher. Figure 3 contains a page from the 4-choice CEs that corresponds to the 3-choice page in Figure 2.

Constructed-response CEs. The constructed-response CEs were formed by simply listing the words, word elements, and word-attack words for the unit on a 5" x 8" card (Figure 4). The teacher administering this type of test simply asks a child to come up and read through the card, while the teacher checks any incorrect responses on a separate score sheet. These tests were limited to fifteen items, or five items per outcome. Most of the instructional units in the program contained fewer than five word elements, and rather than repeat the same letter twice (as was done on the selected response CEs), the five-item block for "Word Elements" was simply not filled up. Thus, the number of items in each of the 14 constructed-response CEs varied from 11 to 15. To prevent teachers from taking undue time to administer this test to a child, special instructions were given not to "instruct" a child when an error occurred, but to go right on and worry about remediation later.

TRYOUT PROCEDURES

Subjects

3- and 4-choice CEs. Half of the children in each of five first-grade classes at a suburban school in a large metropolitan district were randomly assigned to receive either the regular 3-choice CEs or the newly developed 4-choice CEs. To insure that each child always received the proper type of CE, names were written on all 14 tests for each child at the beginning of the year. Since the 3-choice and the 4-choice CEs were



Figure 3. Sample page from one of the 4-choice CEs (corresponds to page from 3-choice CE in Figure 2.)

	I've	fly	eyes	we're
2	° box	blue	by	black
3.	f	у	×	v ,
4	id	eck.	ix	ick
5	six	fix	sit	· mix
6	fix	dig	fig .	fog

Criterion Exercise 8 OUTCOME 1-WORDS 1 cry Directions for items 1-5: "Read number ____." 2 one 3 I've 4 we're blue OUTCOME 2-WORD ELEMENTS ix Directions for items 6-10: "Say the sound of number _ 10 OUTCOME 3-WORD ATTACK 11 mix. Directions for items 11-15: "Read number ____." 12 van 13 six 14 vast fix

identical except for the number of distractors (i.e., the answers to be marked were the same for either type of test), the teacher could administer both types at once using the same set of directions. Appendix A contains written procedures given the teachers for administering both the 3-choice and the 4-choice CEs.

Constructed-response CEs. All of the children in five classes in two suburban schools at another district were assigned to receive the constructed-response CEs during the entire school year. Previous tryouts of the reading program indicated that the achievement levels obtained at these two schools were fairly comparable to those of the school using the 3-choice and 4-choice CEs. Special administration and scoring procedures given teachers for the constructed-response CEs are contained in Appendix B.

Data Sources

In addition to the scores on the three types of CEs during the year, the primary data source for the study were the scores on a 56-item, constructed-response mastery test individually administered by SWRL personnel in all ten classes at the end of the year. (See Appendix C.) The test was administered to a random sample of eight pupils in each of the five constructed-response classes and a random sample of 16 pupils (eight from the 3-choice CE group and eight from the 4-choice CE group) in each of the five classes using both the 3-choice and the 4-choice CEs. This produced a total of 40 posttest scores for each of the three types of tests.

Class Record Sheets containing scores on the Criterion Exercises were also collected from teachers throughout the year. In addition, the completed test booklets themselves were collected from the five classes utilizing the 3-choice and 4-choice, selected-response CEs. Complete posttest and CE data were obtained for 40 children in the 3-choice CE group, 36 children in the 4-choice CE group and 33 children in the constructed-response CE group.

Mid-year meetings were held with teachers in all classes to obtain their comments, criticisms and suggestions regarding the three types of tests.

RESULTS

Pupil Performance Data

Table 1 indicates the correlation between average CE scores and posttest scores for each cf the three groups. It may be seen that there was a high, positive relationship between CE scores and the posttest scores for all three groups. The question still remains, however, which test gives the teacher the most valid estimate of how children will perform at the end of the year?

Table 2 shows the mean percentage scores by outcome on the CEs and on the end-of-year posttest for the pupils in each of the three groups receiving the three types of tests. It may be seen that the average CE scores for the 3-choice CE Ss and the 4-choice CE Ss were 98.2% and 94.0%, respectively. This represents about a one-item difference on the 24-item tests. When it is noted, however, that a child need miss only one item on an outcome to be assigned a remedial Practice Exercise,



Table 1

CORRELATION COEFFICIENTS (PRODUCT MOMENT) OF RELATIONSHIP BETWEEN AVERAGE CE SCORES AND POSTTEST SCORES FOR EACH TYPE OF CE

	3-choice	4-choice	Constructed-
	CEs	CEs	Response CEs
	(n=40)	(n=36)	(n=33)
Correlation coefficient between average CE score and posttest score	.78*	83*	.78*

^{*} p < .01

Table 2

MEANS AND STANDARD DEVIATIONS ON CES AND POSTTEST:
FOR EACH TYPE OF CE

· · · · · · · · · · · · · · · · · · ·	3-choice CE _S (n=40) 文 % sd		4-choice CEs (n=36)		Constructed- Response CEs (n=33) X % sd	
Mean CE score (%)	98.2	4.6	94,0	7.2	89.8	12.3
Mean Posttest Score (%)	86.1	16.0	86.3	15.1	80 . 6	18.7

this one-point difference means that about one more Practice Exercise per child, or about 30 additional Practice Exercises per class, is generated by the 4-choice CEs than by the 3-choice CEs. The mean CE score for the constructed-response CE Ss was 89.8%, but it is hard to compare this group to the 3-choice and 4-choice CE groups, since the constructed-response CE Ss came from a different population than the Ss receiving either 3-choice or 4-choice CEs on a random basis.

The posttest scores in Table 2 reveal very little difference between the 3-choice CE and the 4-choice CE groups, and there is little reason to expect much difference. The lower score for the constructed response CE group probably reflects the fact that these Ss were drawn from a different population than the other two groups.

Figures 5, 6 and 7 are scattergrams showing the relationship of CE scores to posttest scores for the 3 choice CE group, the 4-choice CE group and the constructed-response CE group, respectively. From the scattergram for the 3-choice CE Ss (Figure 5) it may be seen that 11 of the 40 Ss averaged higher than 90% on the CEs, yet scored less than the mastery criterion of 80% on the posttest. Thus, for 28% of the pupils in the 3-choice CE group, the teacher was receiving indication during the year that the children were performing quite well (over 90%), yet those children failed to reach mastery on the end-of-year test. With the other two types of Criterion Exercises, however, this did not happen. For the 4-choice CE group (Figure 6) and the constructed-response CE group (Figure 7), it may be seen that only two Ss in either group (6%) averaged over 90% on the CEs, yet scored below the mastery level of 80% on the posttest.

Figure 5. Scattergram showing relationship between the average CE score and the posttest score for 40 first-grade pupils receiving the 3-choice CEs.

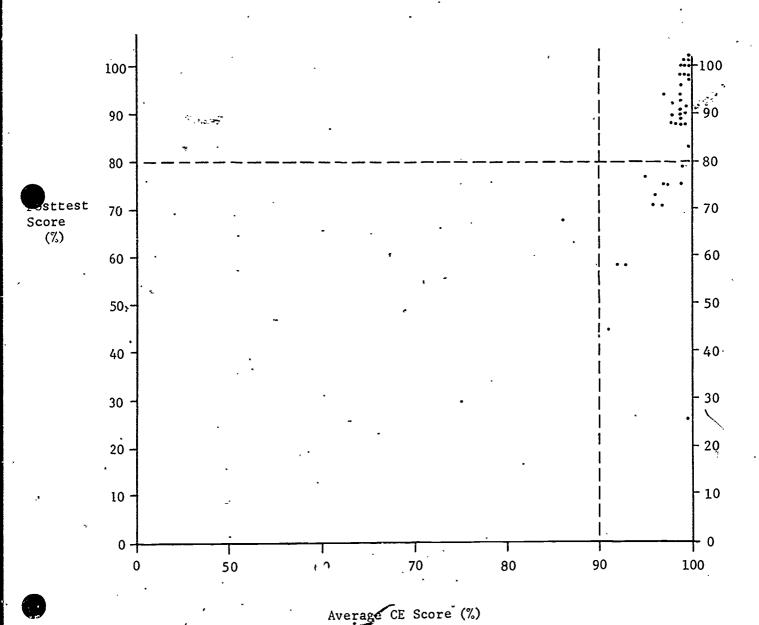


Figure 6. Scattergram showing relationship between the average CE score and the posttest score for 36 first-grade pupils receing the 4-choice CEs.

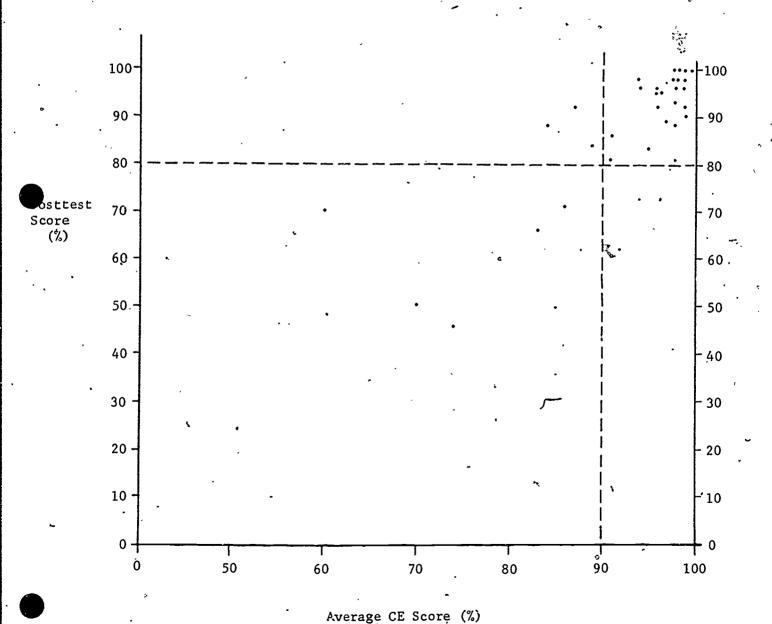
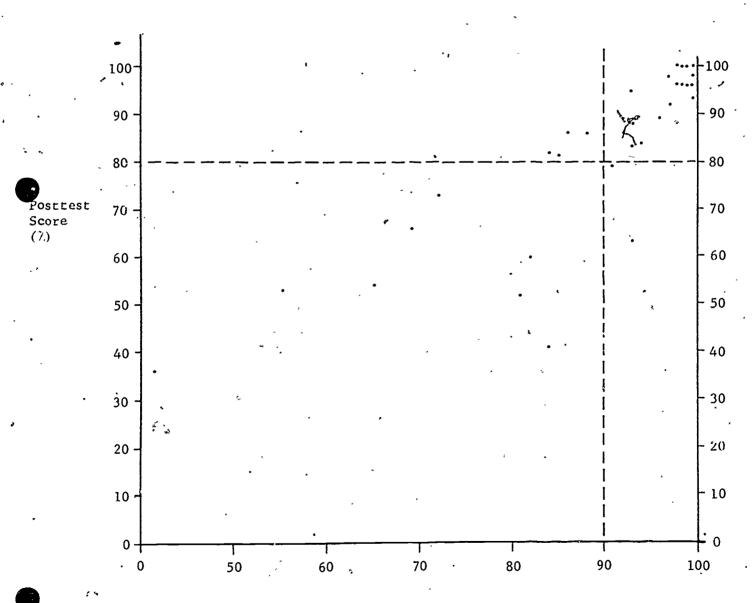


Figure 7. Scattergram showing the relationship between the average CE score and the posttest score for 33 first-grade pupils receiving the contracted-response CEs.

ع



Average CE Score (%)

Results of Teacher Meetings

At a meeting of the five teacher utilizing the 3-choice and the 4-choice CEs, the teachers indicated a preference for the 4-choice CEs. From scoring the CEs before returning them to SWRL, it seemed apparent to these teachers that the 4-choice CEs generated lower scores and was a more valid measure of reading ability.

All five teachers utilizing the constructed-response CEs indicated that they preferred these individually-administered tests over the group-administered, selected-response tests. (These teachers were familiar with the latter type of test in that they had used selected-response CEs the previous year with the SWRL Reading Program.) Among the reasons mentioned for this preference of constructed-response CEs over selected-response CEs were the following:

- · easier to administer
- takes less time (less than one minute per child, or about half the time required to administer, score and record the selected-response CEs to a group of 10 or 12 children)
- eliminates copying
- gives the teacher better knowledge of an individual child's skill level (because the constructed-response CEs are more difficult and because of the one-on-one administration)
- makes the child feel good when he receives the individual attention of the teacher

Each of 10 first-grade teachers involved in the study grouped their classes into three groups of reading instruction.



DISCUSSION

The regular 3-choice CEs normally used in the SWRL First-Grade Reading Program produced a mean score of 98.2% across five classrooms. Twenty-eight percent of the children receiving this type of test averaged above 90%, yet scored below the mastery criterion of 80% on an end-of-year test while averaging above 90% on the CEs during the year. These data, combined with favorable teacher reactions, indicate that either the 4-choice CEs or the constructed-response CEs would be more appropriate than the 3-choice CEs for the SWRL Reading Programs.

The highly positive reactions of the teachers utilizing the constructed-response CEs were somewhat unexpected. It had been felt that a teacher with 30 students would find this type of test too time consuming to administer on a fairly frequent basis. As it turned out, however, each teacher had grouped the children into three groups for reading instruction. Thus, a teacher never had to administer any one CE to more than one group at a time (anywhere from six to twelve children). Had the classes not been grouped, the problem of time may have come up, since the teacher would have to have administered the same test to 30 children at once on an individual basis.

Thus, when teachers group, as is usually done in first-grade and in higher grades, the constructed-response CEs would probably be most appropriate in that this type of test actually has the children read, rather than select, words and sounds. In kindergarten, however, teachers normally have not grouped for instruction and have kept the entire class



together during the year. In this situation, the group-administered, 4-choice CE would probably be most efficient.

In summary, the results of the study indicate that the groupadministered, 4-choice tests would be most appropriate in SWRL Reading
Programs at the kindergarten level and the individually-administered,
constructed-response CEs would be most appropriate at first-grade or
above. The currently used 3-choice CEs do not provide an accurate
indication of achievement for many of the children and their continued
use is not recommended.



APPENDIX .A

Teacher Procedures for 3-choice and 4-choice CEs

Criterion Exercises

Administration

- · Arrange seating so that the children cannot see each other's answers.
- · Be sure each child's name is on his booklet.
- Read each item exactly as printed in <u>Directions</u> for the exercise (Direction Cards in Program File Box.)
- Help children find the correct row or page only if necessary and only for the first two exercises.
- · Do not provide hints or clues to the right answer.
- · Do not give the correct answer after children have marked the item.

Scoring

- · Enter names on Class Record Sheet.
- · Score each booklet using Scoring Key from Directions for the exercise.
- While scoring, make a checkmark beside each item not marked or marked incorrectly.
- After all booklets are marked, total the first child's score for Rows 1 and 2, Rows 3 and 4, and Rows 5 and 6. Enter his score for Rows 1 and 2 in Column 1 in the Class Record Sheet, Rows 3 and 4 in Column 2, and Rows 5 and 6 in Column 3.
- Enter the child's total score for the three outcomes in the "Total Correct" column.



Teacher Procedures for Constructed-response CEs

PROCEDURES FOR INDIVIDUALLY ADMINISTERED CRITERION EXERCISES

The procedures below are for use with the individually administered Criterion Exercises. They replace the procedures for group administered Criterion Exercises listed on page 51 of the Teacher's Manual.

Administration and Scoring¹

- · Place the child so that he won't be distracted by the other children.
- · Locate or write his name on the Criterion Exercise Record Sheet.
- · Place the Criterion Exercise in front of the child so he can see it easily.
- · Read the directions exactly as they are printed for each of the 15 items. 2
- For each item not answered or answered incorrectly, place a check in the appropriate outcome column for the Criterion Exercise being tested. If the child reads all the items correctly, place a check (/) in the 100% column.

1			Cri	ter	<u> </u>	<u> </u>	$\overline{}$
		•	1-				- 5
	~	Exercise 1					
_	Student		Outcome'			0	5
			11_	_ 2	1 3	100	_5
	1. Bobby Jackso	n	VV		1		?
ı	2. Nancy Bennet	t				/	7.
ĺ		٠	سبننا				ì

For example, from the Criterion Exercise Record Sheet above, it may be seen that Bobby Jackson missed two items on Section 1 (Outcome 1: Words--Items 1-5) of the Criterion Exercise for Unit 1, none on Section 2 (Outcome 2: Word Elements--Items 6-10) and one on Section 3 (Outcome 3: Word Attack--Items 11-15). Nancy Bennett, however, read all the items correctly.

- · Help the child find the correct number only if necessary.
- · Do not confirm correct answers.
- · Do not provide hints or clues.
- · Do not give the correct answer if the child is incorrect or does not answer.

Assigning Practice Exercises

Each child with one or more checks in an outcome column on the Criterion Exercise Record Sheet should receive the Practice Exercise corresponding to that outcome. Outcomes 1 (Words), 2 (Word Elements), and 3 (Word Attack) correspond to Practice Exercises "a," "b," and "c" respectively. Procedures for administering the Practice Exercises are listed on page 52 of the Teacher's Manual.

¹It should take only about one minute per child to individually administer the exercise.

²For some units there are less than 15 items, but this should make no difference in the procedures.

RECORD SHEET CRITERION EXERG

Di Sor each incorrect response, place a check in the appropriate outcome box. Check 100% if none are incorrect

Teacher Mrs. Delamere School Fern

Criterion Exercise 7 Outcomes 1 4 2001 Criterion Exercise 6 Outcomes C) ್. 001 Criterion Exercise 5 , ش Outcomes C: **ಬ001** Criterion Exercise 4 Outcomes ್ಲ00≀ Criterion Exercise 3 Outcomes **ಬ00**1 Criterion Exercise 2 Outcomes C.S %00I Criterion Exercise 1 က Outcomes C) Del Palacio, Graciela Carlson, Frederick Thompson, Kimberly 4. Ellingson, Marcus 7. Beedle, Stephanie Harrison, Noelene Anderson, Ronnie Bennett, Robert Schmidt, Jeffrey Sanders, Deborah Steward, Timothy Borland, Kelley 9. Conry, Kimberly Hughes, Michael Moeder, Roxanne Smith, Phillip Infante, Frank Kaiser, Steven Pulliam, Nancy Wambold, Randy Hoien, Melissa Ernst, Crystal Mintz, Mirian Tyler, Kevin' Bauer, Julia Skeen, David Mesa, Vivian Ivers, Cheri Holt, Devin Wade, Susan Pfau, Jack

œ ō

က်

.α

ĸ,

က်

Name

APPENDIX

rogram Posttest SWRL First-Grade Readin

1970-71* SYCSP POSTTEST:

Units 1-14

· School

District,

1 + 1 + Word Attack II*** + 1 + + Outcome 3: "Read this word," fished brave deėd fade role hike West cap lap dom lid 8. . Max Teacher 12. 6 10. 11. Word Attack I** + + Outcome 3: "Read this word." pulled bring sack neck duck vest cave hide bone pilllog tin 12. 10. 11. 'Say the sound of this + 1 + 1 + 1 + Word Elements (these) letter(s)." Outcome 2: ube ime est ate ack do ငှာ 10. 11. + + + + + + + ı + 1 + + + + Outcome 1: "Read this word." Words children outside apple green home poog jump cake cat bed gnq six Date 10. 11. 12.

Total Correct

CORRECT RESPONSES GRAND TOTAL

+

Mort

14.

+

fort

14.

• +

ort

14.

+

story

14.

bird

13.

Total Correct

13.

Fotal Correct

sha.

13.

+

part

13.

Total Correct

^{*}Item number corresponds to the program unit number from which content was selected.

^{***}These words were not seen or practiced in the program, but are composed of Word Elements taught under Outcome 2.